CMPSCI 611: Advanced Algorithms

Lecture 19: Weighted Set Cover and TSP

Andrew McGregor

Last Compiled: April 16, 2024

Outline

Approximation Algorithms Recap

Weighted Set-Cover

Metric TSP 2-approx

Approximation Ratios

Definition

The *performance ratio* of an algorithm is

$$\max_{x:|x|=n} \frac{C_{alg}(x)}{C_{opt}(x)} \quad \text{for a minimization problem}$$

$$\max_{x:|x|=n} \frac{C_{opt}(x)}{C_{alg}(x)} \quad \text{for a maximization problem}$$

where $C_{alg}(x)$ is the value of the algorithm solution on input x and $C_{opt}(x)$ is the value of the optimal solution on input x.

Outline

Approximation Algorithms Recap

Weighted Set-Cover

Metric TSP 2-approx

Set-Cover

Problem:

- Input: A collection $C = \{S_1, S_2, \ldots, S_m\}$ of subsets of $U = \cup_{S \in C} S$ and weights $w : C \to \mathbb{R}^+$
- Output: Find $C' \subset C$ such that

$$U = \cup_{S \in C'} S$$

that minimizes $\sum_{S \in C'} w_S$.

Greedy Set-Cover Algorithm

Algorithm

1. *Let* $R \leftarrow U$, $C' \leftarrow \emptyset$
2. *While* $R \neq \emptyset$*:*
   2.1 *Pick* $S \in \{S_1, \ldots, S_m\}$ *be the set minimizing* $w_S / |S \cap R|$
   2.2 $R \leftarrow R - S$ *and* $C' \leftarrow C' \cup \{S\}$
3. *Return* C'

Theorem

Approx ratio is $\frac{1}{d^*} + \frac{1}{d^*-1} + \ldots + \frac{1}{1} \approx \ln d^*$ *where* $d^* = \max_i |S_i|$

Definition

When S is chosen, say $e \in S \cap R$ is covered at cost $c_e = \frac{w_S}{|S \cap R|}$. Note

$$\sum_{S \in C'} w_S = \sum_{e \in U} c_e \ .$$

Ex: If $S_1 = \{1, 2\}, S_2 = \{1, 2, 3\}, S_3 = \{3, 4\}$ where $w_{S_1} = 4$, $w_{S_2} = 7$, $w_{S_3} = 20$ then $c_1 = c_2 = 2, c_3 = 7$, and $c_4 = 20$.

Analysis 1/2

Claim

For all $S \in C$, $\sum_{e \in S} c_e \leq H(|S|) \cdot w_S$. Note that S may or may not be one of the sets chosen by the greedy algorithm.

Proof.

- Suppose $S = \{e_1, \dots, e_d\}$ be ordered according to order in which elements are covered by the greedy algorithm (break ties arbitrarily).
- Suppose S' is chosen to cover e_j. Because algorithm is greedy,

$$c_{e_j} = w_{S'}/|S' \cap R| \leq w_S/|S \cap R|$$

- Before e_j was covered $e_{j+1}, \dots, e_d$ were also uncovered,

$$|S \cap R| \geq (d - j + 1)$$

- Therefore, $\sum_{j=1}^{d} c_{e_j} \leq \sum_{j=1}^{d} \frac{w_S}{d-j+1} = \frac{w_S}{d} + \frac{w_S}{d-1} + \dots + \frac{w_S}{1}$

□

Analysis 2/2

- Let the optimal solution C_{OPT} have cost w_{OPT}.
- Claim: For each $S \in C_{\text{OPT}}$,

$$w_S \geq \sum_{e \in S} \frac{c_e}{H(|S|)}$$

where $H(|S|) = \frac{1}{|S|} + \frac{1}{|S|-1} + \ldots + \frac{1}{1}$.

- Then,

$$w_{\text{OPT}} \geq \sum_{S \in C_{\text{OPT}}} \sum_{e \in S} \frac{c_e}{H(|S|)} \geq \frac{1}{H(d^*)} \sum_{S \in C_{\text{OPT}}} \sum_{e \in S} c_e \geq \frac{1}{H(d^*)} \sum_{e \in U} c_e$$

- But $\sum_{e \in U} c_e = \sum_{S \in C'} w_S$ and so $w_{\text{OPT}} \geq \frac{1}{H(d^*)} \sum_{S \in C'} w_S$

Outline

Approximation Algorithms Recap

Weighted Set-Cover

Metric TSP 2-approx

Metric Traveling Salesperson Problem

- Input: Weighted complete graph $G = (V, E)$ with positive weights such that for edges $e = (u, v), e' = (v, w)$, and $e'' = (u, w)$

$$w_e + w_{e'} \geq w_{e''}$$

- Goal: Find the tour (a path that visits every node exactly once and returns to starting point) of minimum total weight.

Metric TSP Approximation Algorithm

Algorithm

1. *Compute minimum spanning tree T_{mst} of G*
2. *Consider a “pseudo-tour” that walks around T_{mst}*
3. *Create tour from pseudo-tour by skipping pre-visited nodes*

Theorem

The algorithm is a 2-approximation.

Proof.

- Cost of pseudo-tour is twice cost of T_{mst}
- Cost of tour found is at most cost of pseudo-tour:

$$\text{cost}(\text{tour found}) \leq \text{cost}(\text{pseudo tour}) = 2 \cdot \text{cost}(T_{mst})$$

- Cost of T_{mst} is at most cost of optimal tour since removing an edge in an optimal tour gives a spanning tree:

$$\text{cost}(T_{mst}) \leq \text{cost}(\text{optimal tour})$$

□